Tacit Contracts for Wheelchairs

Daniel Couto Vale
SFB/TR8 Spatial Cognition
University of Bremen
danielvale@uni-bremen.de

Vivien Mast
SFB/TR8 Spatial Cognition
University of Bremen
viv@tzi.de

Abstract

In this paper, we propose a novel approach to infer dialogue acts using the notion of *tacit contracts*. We describe the interpersonal linguistic features that our analysis grammar can identify in uttered texts and present an inference procedure that strictly separates the semantic and pragmatic steps of utterance understanding, thereby meeting a higher degree of modularity, a prerequisite for extending robot functionality.

Keywords: Dialogue System; Dialogue Act; Attitude; Stance

1 Introduction

John is reading "Merlin", when the door bell rings. He cannot walk, but his intelligent wheelchair Rolland is nearby. He says: "Rolland, I need to open the door. Can you take me there?" Rolland responds "Sure, I'm coming!", comes to him, waits for him to sit comfortably, and then says "Let's go!" before driving him to the door side where John is able to reach the door handle.

So seamless are the interactions in our Wizard-of-Oz experiment (Anastasiou et al., 2012; Vale and Mast, 2012b) yet so difficult for an intelligent wheelchair. How is it to know that "I need to open the door" and "Can you take me there?" should not be understood separately as a statement and a question but together as a command to move towards the door for enabling the user to open it?

Each utterance is an action that affects the interactive situation. Not only does it construe events, but it also constitutes exchanges between interactants such as stating the speaker's need of performing an action and asking about the listener's capability of providing a service.

All speaking robots need some method of coping with this dual character of situated utterances. A frequent approach is Dialogue Act Detection, a family of statistical methods trained on human-annotated corpora (Allen and Core, 1997; Jurafsky et al., 1997; Jurafsky et al., 1997; Jekat et al., 1995). An alternative approach is Plan Recognition, which consists of using a planner having linguistic meaning and a domain model as inputs. We depart from this tradition by proposing a contractual approach in which semantic and pragmatic aspects of understanding are symbolically explored in separate steps of inference.

The main rationale for not pursuing the detection of dialogue acts as patterns in the uttered text is that the intended effect of an utterance is not to be found in the wording (Marcu, 1997); and the rationale for not taking construed events directly as plan steps is that the functional roles of interactants in cooperative work are decisive for interpreting attitudes.

In this paper, we present an automatic semantic analysis of the interpersonal features of linguistic units and propose a compatible three-step procedure consisting of transformation, contextualization and inference to enable an intelligent wheelchair to understand implied dialogue acts.

In the following two sections, we describe prior approaches to understanding interpersonal meaning and discuss their relevance for our approach. Then we introduce a classification of interpersonal linguistic features and explain how we re-express these implicit features of language explicitly in a standardised format. Finally, we describe the procedure used by the pragmatic module of our wheelchair in order to contextualise utterance meaning and infer implied dialogue acts.

2 Dialogue Act Detection

Dialogue act detection is the most frequently used approach for dealing with the interpersonal aspect of dialogue. In this section, we review two frameworks for dialogue act detection: the annotation standard DAMSL (Allen and Core, 1997) and the

Proceedings of the SIGDIAL 2013 Conference, pages 294–303,
Metz, France, 22-24 August 2013.

dialogue act component of the successful application Verbmobil (Jekat et al., 1995).

2.1 DAMSL and Derivates

The tagging system of Dialogue Act Markup in Several Layers (DAMSL) is a tag system for speaker's intention. It uses binary decision trees for tagging utterances with up to four attributes (layers) of the speaker's intention: communicative status, information level, backward- and forward-looking functions. DAMSL has been thoroughly tested for annotation (Core and Allen, 1997; Jurafsky et al., 1997; Ivanovic, 2005; Stolcke et al., 2000) with inter-annotator agreement reaching approx. 70-85%. Recent attempts to automatise DAMSL dialogue act detection using statistical methods (Core, 1997; Rosset and Lamel, 2004; Rangarajan Sridhar et al., 2007; Rosset et al., 2008; Rangarajan Sridhar et al., 2009) reach similar accuracy.

However, high accuracy scores need to be relativized, as precision and recall may be very low for most tags— Rangarajan Sridhar et al. (2009) report $<1\%$ for all but the two most frequent tags. Moreover, agreement rates tell nothing about the severity of mistagging for application usage.

The second and most compromising issue lies in the annotation scheme itself. Context-dependent decision trees turn utterance tagging into guess work, since utterances map differently to world models in different situations. For instance, there is danger "of annotators confusing surface form with [contextual] speaker intent, for instance labelling an *info-request* in the form of a statement as an *other-statement*" (Stent, 2000).

The third issue concerns applicability. The DAMSL research community has built annotated corpora and automated dialogue act detection. Only lately, there has been research on automatic learning of dialogue act flow patterns on large manually annotated corpora of dialogue. Whether these dialogue act flows will be usable in real applied systems is yet to be determined.

2.2 Verbmobil

A different approach to Dialogue Act Detection is followed by Verbmobil, a successful applied dialogue system for travel booking. It uses 55 types of dialogue act, tailored to the particular application domain of travel booking, e.g. *Request-Suggest-Duration*. Classification of utterances is achieved by detecting keywords and syntactic patterns in the word sequence of the utterance and matching them against keyword and pattern lists which are typical for each dialogue act type. Ambiguity is solved by using a context-based preference order learnt from a large annotated corpus.

This approach works with a caveat: embedding domain content like stay duration in dialogue act types may cause an explosion of categories for less restricted domains and, while easily recognisable, such tailored categories are domain specific. Therefore, they are not reusable when creating an application for a new domain.

2.3 Detection Trade-Off

In short, we argue that statistical methods of dialogue act detection do not scale. This approach always leads to a trade-off between suboptimal inter-annotator agreement as in DAMSL or lack of reusability as in Verbmobil.

The reason for this trade-off is that two issues of different natures are tackled at the same time: semantics and pragmatics. The lack of a stratified linguistic theory with semantic and pragmatic steps behind these classifications is the cause of a bad fit between categories and grammatical structure in the case of DAMSL. This shortcoming can be partially overcome through the usage of tailored categories at the expense of large annotated training corpora and a low reusability.

Tailoring is particularly expensive when experimental data is not easily obtainable as for human-wheelchair interaction. Therefore we need a different approach that separates text analysis from utterance contextualisation (see Section 4).

3 Belief-Desire-Intention Approach

The formal theory of rational interaction (FTRI) is a plan-based approach to dialogue management that divides user mental representations into beliefs, desires and intentions (BDI). The dialogue manager keeps track of which planned tasks are feasible for, assigned to and/or completed by particular agents (Sadek, 1992; Sadek, 1994; Sadek et al., 1996; Sadek et al., 1997). Logical inferences are made with respect to interactant's mental states in order to plan the next verbal action of the dialogue system. The interpersonal features of the linguistic model are very simple. User utterances are classified into one of 3 categories of syntactic patterns (directive, interrogative, or affirmative),

and are used in combination with spotted verbs in order to determine the beliefs, desires, and intentions of the user. Lists of implications are used in order to infer actions from certain combinations of intentions and beliefs. An example for such an implication is: If a user *u* intends (*I*) to have an action *a DONE*, *u* intends (*I*) her/his utterance to have the same rational effect *RE* as performing *a* by her/himself; in other words, by informing the system about her or his intention to have a task performed, the user delegates the performance of this task to the system. The formal theory of rational interaction along with similar theories have received strong criticism. They lack a formalization of linguistic meaning (structural meaning) capable of encapsulating the richness and flexibility of linguistic systems. Moreover, "[BDI p]lan-based approaches [to dialogue management] are also criticised as being more opaque, especially given the large amount of procedural processing and lack of a well-founded semantics for plan-related operations" (Traum et al., 1999). Improving upon BDI plan-based approaches, Traum's work takes into account dialogue acts (Traum and Hinkelman, 1992) and obligations (Traum and Allen, 1994). Reflecting Traum et al.'s critique of the BDI approach, the present work can be understood as a step towards the theoretical conceptualization of some of BDI's opaque operations within an information-state approach to dialogue management.

4 Contract-Supported Interaction

Our work is supported by automatic functional text analysis (parsing) with Combinatory Categorial Grammar (CCG) (Steedman and Baldridge, 2011) using Systemic Functional Theory (Halliday and Matthiessen, 2004). This enables us to detect personal stances and attitude automatically in the syntactic structure of the utterance (see Section 4.1). Based on these detected concepts, we generate a standardised typed feature structure, which captures the commonalities of different utterance types with respect to the implied expectations from the addressee. Rather than hypothesizing about the user's mental states, we are able to base our interpretation solely on what is linguistically expressed as required from the addressee. Our usage of inference is somewhat similar to the one proposed by FTRI; however, with the formalised concept of tacit contracts—further formalizing the update operations in the dialogue system of Matheson et al. (2000)., we gain situational flexibility which is of great value, because tacit contracts are not universally valid, but depend on the roles of the interactants in a given situation.

4.1 Interpersonal Upper Model

On the semantic level, we adopt the most comprehensive linguistic description of the interpersonal component of human languages, which is found in the Systemic Functional Grammar of English (Halliday and Matthiessen, 2004). We created an ontology of linguistic units, the *Interpersonal Upper Model*, covering all interpersonal features of Systemic Functional Theory with minor adjustments and extensions. Using this classification of linguistic units, we implemented a Combinatory Categorial Grammar for German to parse a corpus collected in a Wizard-of-Oz experiment where users gave commands to an intelligent wheelchair in order to perform simple domestic tasks like washing their hands or opening the door (Anastasiou et al., 2012; Vale and Mast, 2012b; Vale and Mast, 2012a).

For an intelligent wheelchair to understand the utterances of the user, first it must cope with the various ways in which interpersonal meaning is expressed in language. For example, by uttering either of the clauses "I want you to leave." or "Leave!", the speaker commands the addressee to perform the action of leaving. Although they share this interpersonal function, the first makes the command explicit by referring to the requirer and performer of the service while the second leaves it implicit in the structure of the clause.

Halliday and Matthiessen (2004) call such expression pairs, where different wordings represent the same interpersonal meaning, interpersonally *agnate expressions*. In their work, *grammatical metaphor* is defined as the process whereby concepts which are usually implicit in the structure of clauses are re-expressed with more explicit referential representations.

It must be noted that this analysis relies strictly on utterance semantics, i.e. the information that can be gained from automatically analyzing the utterance alone, without relying on linguistic or situational context. By relying on parsing rather than string-level methods such as POS-tagging, keyword spotting and statistical utterance classification, we have the advantage of retaining the

rich information contained in the structure of the utterance. For our approach, we rely on parsing with Categorial Combinatory Grammar (CCG) (Steedman and Baldridge, 2011) based on Systemic Functional Theory. This methodology provides us with a systematic, theory-based way of retrieving features of the linguistic structure of an utterance that are relevant for human-computer interaction. When parsing with a functional grammar, syntactic units are classified according to their function. Therefore, the segmentation of the utterance into constituents is based on the compositionality of semantic units.

In the remainder of this section, we will explain in detail the two main characters of interpersonal meaning, *attitude* and *stance* and how they are recognised in linguistic structure. In the following section, we will proceed to demonstrate how we turn the concept of grammatical metaphors into a method for representing interpersonal linguistic features explicitly in a standardised manner.

4.1.1 Attitude

Attitudes (or direct speech acts) specify the kind of thing negotiated: a mercative attitude indicates an exchange of goods ("A beer, please!"), an imperative attitude an exchange of services ("Please take me to the kitchen!"), and a declarative attitude an exchange of information ("Is the door closed?"). They also specify the orientation of the exchange between the interactants: whether the speaker is offering something to the addressee ("Your beer!" – offertive attitude) or demanding something from them ("A beer please!" – mandative).

By classifying attitudes in these two dimensions, we have a clear separation of exchange orientation (speaker to addressee or vice-versa) and exchange stock (good, service, or information). The combination of these options yielding six different attitudes[1], as shown in Table 1.

Table 1: Orientation × Stock → Attitude

	Orientation × Stock	Attitude
demand info	mandative × declarative	interrogative
offer info	offertive × declarative	affirmative
demand service	mandative × imperative	directive
offer service	offertive × imperative	preemptive
demand good	mandative × mercative	questive
offer good	offertive × mercative	donative

[1]The full ontology contains more distinctions that are ignored here for the sake of brevity.

As Example (1) shows, a mercative attitude (exchange of goods) is usually expressed by noun groups with modifiers such as "please". There is no constituent for Process nor Subject. Imperative attitudes (exchange of services) are usually expressed by predicates, that is, they have no Subject constituent as shown in Example (2). Finally, declarative attitudes (exchange of information) are usually expressed by full predications [2], as in Examples (3) and (4).

(1) "A beer, please!" (mercative)
(2) "Please take me to the kitchen!" (imperative)
(3) "The door is closed." (declarative)
(4) "Is the door closed?" (declarative)

Because there is a mapping between the syntactic level of an utterance structure and the kind of stock being exchanged (goods, services or information), we can automatically detect which attitude each utterance has.

4.1.2 Stance

Stance (or modality) "construe[s] the region of cognitive uncertainty that lies between 'yes' and 'no"' (Halliday and Matthiessen, 2004). There are two primary kinds of stance: *control* determines whether someone wants something (inclination, e.g. "is keen to", "wants") or is wanted for something (regulation, e.g. "is supposed to", "must"), and *conviction* determines how likely something is (likelihood, e.g. "it's likely to rain", "it's definitely not going to rain"), or how often it occurs (usuality, e.g. "It often rains in summer.", "it never rains in the desert."). Because Systemic Functional Theory works by delimiting semantically classified syntactic units based on possible semantic oppositions, combinatory categorial parsing of expressions is straight forward.

4.2 Grammatical Metaphor

As discussed in the previous section, attitude has an orientation from the speaker to the addressee or vice versa (offering or demanding). The service requirer and/or provider are not explicitly mentioned, but determined by the syntactic structure used and the roles of the interactants in the dialogue, speaker and addressee. Halliday and Matthiessen (2004) call this *interpersonal orientation* . Stance provides a linguistic tool to explicitly express the source and target of orientation, detaching them from the interactional situation.

[2]i.e. association between a subject and a predicate

For instance, "Leave!" is a service demand with an *interpersonal orientation* from the addressee to the speaker. If one rephrases this with "must" as in "you must leave." or "he must leave.", one obtains a *personal stance*, that is, an orientation from the speaker to the provider of the service which is explicitly expressed by a reference such as "you" or "he". By rephrasing the utterance again with "require" as in "you are required by me" or "you are required by law", one obtains an *impersonal stance*, that is, both requirer and service performer are referred to explicitly and not assumed from the orientation of the linguistic exchange.

It is possible to express the same interpersonal meaning with an *impersonal orientation* as with an *interpersonal orientation*. For example, "You are required by me to leave.", just like "Leave!", takes the speaker as the requirer and the addressee as the performer of the required action of leaving, therefore these two expressions are *agnate*, making the first a grammatical metaphor of the second.

Table 2: Possible orientations.

Interpers.	Personal	Impersonal
leave	you must leave he must leave	you are required by me to leave he is required by me to leave you are required by law to leave he is required by law to leave

4.2.1 Addressee-centered perspective

Each attitude brings about a *required response* from the addressee: offertive attitudes, by offering a stock, pose a requirement to receive this stock and mandative attitudes, by demanding a stock, pose a requirement to give one. These required responses can be expressed explicitly in more metaphorical agnate expressions. For example, the attitude of offering goods (offertive $\times$ mercative $\rightarrow$ donative) is represented by the process "take" in agnate expressions with the addressee as the subject as in "Take some cookies.". Table 3 shows the mapping of all 6 main attitudes onto their corresponding requirements from the addressee.

With mappings from the less metaphorical expressions to more metaphorical ones, the wheelchair can construe a standardised semantic representation to work with. This explicitation method enables us to capture the semantic commonalities of a broad variety of different linguistic expressions. Examples (5) and (6) show two different utterances whose standardised representations are highly similar. Example (5) is an information offer, re-expressed as a requirement to know a given information. Agnately, Example (6) is a service demand, re-expressed as a requirement to perform the service of *being aware* of the same information, a particular way of knowing it[3].

Table 3: Mapping of attitudes onto requirements from the addressee

Attitude	Required Reaction	Process
donative	receive goods	take
questive	give goods	hand
preemptive	receive services	assign
directive	give services	perform
affirmative	receive information	know
interrogative	give information	say

(5) "it's snowing"

LINGUISTIC MEANING:

```
Speaker offers to Addressee information that
     it's snowing
```

ADDRESSEE-CENTERED MEANING:

```
Speaker requires Addressee
     to know that
          it's snowing
```

(6) "be aware that it's snowing"

LINGUISTIC MEANING:

```
Speaker demands from Addressee service
     of being aware that
          it's snowing
```

ADDRESSEE-CENTERED MEANING:

```
Speaker requires Addressee
     to perform
          being aware that
               it's snowing
Speaker requires Addressee
     to be aware that
          it's snowing
```

The standardised semantic representation has the advantage that the wheelchair needs to treat requirements in only the most explicit representation when deciding which action it is expected to perform. In the following section, we will explain the concept of *tacit contracts*, and how they are used by our *interpersonal calculus* in order to extract the dialogue act from the user utterance as represented by the addressee-centered semantic representation and the situation model.

4.3 Tacit Contracts

While the addressee-centered semantic treatment enables an intelligent wheelchair to deal with utterances such as (7) and the more metaphorical (8) in a standardised manner independent of the situation, there is a further step of processing needed

[3] As performing an action is the same as acting, in Example (6), "requiring to perform being aware" can be simplified to "requiring to be aware".

in order to deal with the full scope of utterances collected in our usability experiment.

(7) "Take me to the kitchen."

(8) "I want you to take me to the kitchen."

(9) "I must go to the kitchen."

For instance, the wheelchair needs to understand that utterance (9) is, in the dialogue situation, not only an offering of information of a need of the user, but a more polite variant to Examples (7) and (8) (Vale and Mast, 2012a). The meaning the speaker intends to convey goes beyond *what is said*. Grice (1975) called this kind of pragmatic inference *conversational implicature*. They arise from the the understanding of a set of *conversational maxims* which humans can be expected to observe in conversation in combination with features of the interactional situation in which it is uttered. In contrast, *conventional implicatures* arise from the meaning of the uttered sentence and the *maxims of communication*, without any influence from the interactional situation. Récanati (1991) improved the Gricean model of maxims, but for theoretical reasons accepted no linguistic formalism, which makes his model impossible to apply in intelligent wheelchair design.

Relevance Theory (Sperber and Wilson, 1995; Carston, 1998) further develops the concept of inference in a cognitive paradigm by replacing maxims of communication with a balance between the cognitive effort needed to make an inference and its positive cognitive effect under the principle of relevance. Like Récanati, they establish the *linguistic meaning* as the boundary between semantics and pragmatics and divide the inferential process into the two subprocesses *enrichment* (resulting in the *explicatum*) and *deduction* (resulting in the *implicatum*).

As the main aim of this theory is to explain human cognition and not to design artificial intelligence, it is not directly translatable into a method for automation in applied robotics. One problem for automation is the assumption that interactants access and use all kinds of information, as needed. The inherently open nature of this theory makes its operationalization as a general framework impossible. In addition, assessing the *relevance* and cognitive effort of every item of information and process of reasoning makes it computationally too complex for practical applications. Moreover, Relevance Theory is not backed by a grammatical theory, and therefore lacks a comprehensive set of interpersonal linguistic features such as attitudes and stances.

In our approach, we follow the principle of separating meaning into linguistic meaning, explicatum, and implicatum, as proposed by Relevance Theory. Instead of the general effort-effect balance, we propose the concept of *tacit contracts* which operate on the pragmatic deduction step of communication in the Relevance Theory framework. Tacit contracts also differ from Grice's system of conversational maxims, which is not specific enough to distinguish which inferences are expected from particular individuals in their current functional roles.

Rather than general maxims of communication, tacit contracts are specific agreements entered into by two or more parties that establish obligations between those parties. These contracts determine the services that a party is required to perform in given situations. Therefore they determine the services that the speaker can expect from the addressee when he or she causes these situations to happen. For example, a contract such as "your wish is my command" only applies to interactants occupying a given role in the interaction, such as caregiver, waiter, etc., and only for a given set of actions that correspond to this role. If Karl is sitting in a café and says to the waiter "I would like steak for the main course.", the waiter would treat this wish as a command to serve the desired food, because bringing food is part of his tacit contract as a waiter. If Karl were to state "I would like to have your hat.", the waiter would not consider this a command, but a statement, because, although he would be capable to do so, providing the hat is not part of his contract as a waiter.

Politeness, in this perspective, is a manner of obtaining a stock whereby a speaker replaces his or her requirement for an addressee to give out a stock with an exchange of information about the current situation. The new information triggers a tacit contract which then enables the addressee to infer the contractual requirement for the current situation in a separate step of deduction.

4.3.1 Interactional Situation

For inferring implicata it is also important to differentiate two types of *businesses*: *stocks* and *issues*. For instance, in the afore-mentioned table-attending situation, let's suppose Karl had said the same utterance to his friend Hanna "I would

like steak for the main course. Because Hanna cannot give a steak to Karl, the business of this interaction, providing a steak, is an issue and not a stock—they cannot exchange it, but only talk about it. The difference to the hat example above is that the waiter can provide his hat, but is not required to do so by any applicable contract, whereas Hanna may want to provide the steak, but is not able to[4].

By classifying businesses into stocks and issues, it is possible to trim down the inferential process further to avoid undesired implicatures. For instance, a wheelchair should treat the following two utterances differently: 1. "I would like to go to the kitchen" and 2. "I would like to open the door". Taking someone to the kitchen is a stock in this interaction—a service that the wheelchair can perform and that the user can assign to it. Opening the door, on the other hand, is not in the range of the wheelchair's capabilities and therefore an issue.

In the following subsection, we will explain how contracts and business kinds are used in the inferential calculus for generating an implicatum out of the explicatum. Then we will proceed to present the specific contracts relevant for the interaction with an intelligent wheelchair.

4.3.2 Contractual Calculus

Reference resolution and all other situational attachments of meaning are dealt with in the enriching step of the inferential process. Example (10) shows the enrichment of an utterance in the wheelchair scenario.

(10)"I would like to go to the kitchen"

ADDRESSEE-CENTERED MEANING:
Speaker requires **Addressee**
to know that
Speaker *would like*
to go **to the kitchen**
EXPLICATUM:
JOHN requires **ROLLAND**
to know that
JOHN *is keen*
to go to **KITCHEN#1**

After enrichment, the contractual phase is entered. Contracts may be triggered by a requirement towards the wheelchair to say or know. This is then re-interpreted as a polite requirement to give or receive goods, services, or information depending on the contract.

The process for selecting applicable tacit contracts is the following: once a declarative requirement has been detected, the system checks whether the speaker is the *requirer* and the addressee the *provider* of the impersonal stance. If so, for each known contract, it is determined whether the contract applies for the given requirer and provider in their current functional roles. For each applicable contract, the contract script is performed, as will be shown in the following section.

[4]Notice that this reasoning constraint is similar to the Feasibility predicate of FTRI.

Table 4: Surrogation

User: "I need to go to the kitchen."

I	need	to go	to the kitchen
Subject	Finite	–	–
Actor	–	Process	Destination
Medium	–	–	–

Wheelchair: "Ok, I'll take you there."

I	'll	take	**you**	there
Subject	Finite	–	–	–
Actor	–	Process	Action-Goal	Destination
Agent	–	–	**Medium**	–

4.3.3 Wheelchair Contracts

Here we present the contracts needed for understanding the utterances that occurred in our wheelchair-usage corpus. All user utterances in the corpus, except for three cases, can be understood appropriately with the given contracts.

Surrogation is the contract whereby a statement by the speaker of their inclination or obligation to perform an action is interpreted as a demand of a service. For example, if the user puts a bottle on the intelligent wheelchair and tells it "I need to take this bottle to Hannah", the wheelchair should treat this as a command to take the bottle to Hannah, assuming she is close by (similar to the implication in FTRI discussed in section 3).

For a non-affecting action such as "going", the entity that undergoes change as a result of the action (the *Medium*) is the *Actor*. In an affecting action such as "put", on the other hand, the *Medium* is the *Action-Goal* or action target—the thing being put. If a person states that they need to perform an action, the wheelchair needs to perform a service in which it is the *Actor* and which imposes the same result on the *Medium*. As Table 4 shows, this entails substituting a non-affecting action ("go") with an affecting action ("take").

Example (11) shows the performance of a contractual implicature in the deductive process.

(11)"I need to go to the kitchen"

EXPLICATUM:
JOHN requires ROLLAND
to know that

```
          JOHN is required
               to go to the kitchen
IMPLICATUM:
JOHN politely requires ROLLAND
     to take JOHN to the kitchen
```

Supply is a contract whereby *requiring X to say whether X will do* should be interpreted as *requiring X to do*.

Need gleaning is a contract whereby the addressee is required to interpret a question about the availability of a stock as a statement of its need by the speaker. This contract is used together with Surrogation to create polite commands. Example (12) shows the inference of first applying the contract *need gleaning*, interpreting a *requiring X to say whether X can do Y* as a *requiring X to know that Y needs to be done*, and then applying the contract *surrogation*, as described above.

(12)"Can you take me to the kitchen?"

```
EXPLICATUM:
JOHN requires ROLLAND
     to say whether
          ROLLAND can
               take JOHN to the kitchen
IMPLICATUM: NEED GLEANING
JOHN politely requires ROLLAND
     to know that
          JOHN needs
               to be taken to the kitchen
IMPLICATUM: SURROGATION
JOHN requires ROLLAND very politely
     to take JOHN to the kitchen
```

Support is a contract whereby the statement of the speaker's inclination or obligation to perform an action is understood as a command to offer the stock that serves to fulfill the preconditions for performing his or her intended or required action. As Example (13) shows, *requiring X to know that Y is keen to* is interpreted as *requiring X to perform an action that enables Y to*.

(13)"I'd like to open the door!"

```
EXPLICATUM:
JOHN requires ROLLAND
     to know that
          JOHN is keen
               to open the door
IMPLICATUM: SUPPORT
JOHN politely requires ROLLAND
     to take JOHN to a place
          where JOHN can open the door
```

This contract is dependent on the classification of entities by affordances and usage preconditions. A wheelchair can only decide where to take the user who says "I would like to do a mouth wash", if it knows that doing a mouth wash requires the user to be at a certain position in front of a wash basin.

In addition, in order to distinguish whether to apply the contract *support* or the contract *surrogate*, the distinction between *stock* and *issue* is central. If the desired action of the speaker is a *stock*, i.e. a service that can be performed by the wheelchair, the contract *surrogate* should be applied. If it is an *issue*, the contract *support* should be applied instead.

5 Discussion and Outlook

We have presented the main linguistic features of our Enactment Upper Model and shown how to infer dialogue acts by using tacit contracts. With this procedure, we are able to determine automatically which actions the wheelchair is expected to do for most utterances of our corpus. From a theoretical point of view, we proposed a method of deducing implicata by applying contractual scripts that combine a linguistic and a philosophical approach with the strict purpose of automation and, in specific, of controlling an intelligent wheelchair. In doing so, we fill the gap between a linguist's set of lexicogrammatical features with requiring force and a philosopher's set of axioms from which it can be deduced whether the user made a request.

On a robot design perspective, we have spared the text analysis component from creating specific speech acts for a number of clause structures such as "can you..." and "will you..." and so on, which would otherwise be necessary, and spared the dialogue manager from managing a large number of user's dialogue-related intentions and from dealing with the otherwise present ambiguity of contextually interpretable utterances such as "I need to open the door". In addition, our approach enables adjustment for new wheelchair functionality without rewriting the whole text analysis component and allows for an easy addition of new tacit contracts with corresponding scripts.

The approach presented in this paper provides a principled way for inferring dialogue acts that uses the structural information present in the clause and therefore enables high accuracy and reusability both on the semantic and on the pragmatic level. In order to gain a full understanding of the scalability of this approach, further investigation of applicable contracts in different application domains is necessary.

Acknowledgments

We gratefully acknowledge the support of the Deutsche Forschungsgemeinschaft (DFG) through the Collaborative Research Center SFB/TR8 Spatial Cognition. We also thank Dimitra Anastasiou for collaboration in designing and conducting the experiment.

References

James F. Allen and Mark G. Core. 1997. Draft of DAMSL: Dialog Act Markup in Several Layers. retrieved from: http://www.cs.rochester.edu/research/speech/damsl/RevisedManual/, July 16th 2013.

Dimitra Anastasiou, Cui Jian, and Desislava Zhekova. 2012. Speech and Gesture Interaction in an Ambient Assisted Living Lab. In *SMIAE '12 Proceedings of the 1st Workshop on Speech and Multimodal Interaction in Assistive Environments*, pages 18–27, Jeju, Republic of Korea. Association for Computational Linguistics.

Robyn Carston. 1998. The Semantics/Pragmatics Distinction: A View from Relevance Theory. *UCL Working Papers in Linguistics*, 10:303–329.

Mark G. Core and James F. Allen. 1997. Coding Dialogs with the DAMSL Annotation Scheme. Technical report.

Mark G. Core. 1997. Analyzing and Predicting Patterns of DAMSL Utterance Tags. In *Working notes AAAI spring symposium on applying machine learning to discourse processing*, pages 18–24.

Herbert P. Grice. 1975. Logic and Conversation. *Syntax and Semantics 3: Speech arts*, pages 41–58.

Michael A. K. Halliday and Christian M. I. M. Matthiessen. 2004. *Introduction to Systemic Functional Grammar*. Arnold, London, 3 edition.

Edward Ivanovic. 2005. Dialogue Act Tagging for Instant Messaging Chat Sessions. In *Proceedings of the ACL Student Research Workshop*, pages 79–84, Stroudsburg, PA, USA. Association for Computational Linguistics.

Susanne Jekat, Alexandra Klein, Elisabeth Maier, Ilona Maleck, Marion Mast, and J. Joachim Quantz. 1995. Dialogue Acts in VERBMOBIL. Technical report.

Daniel Jurafsky, Elizabeth Shriberg, and Debra Biasca. 1997. Switchboard SWBD-DAMSL Shallow-Discourse-Function Annotation Coders Manual, August.

Daniel Marcu. 1997. Perlocutions: The Achilles' Heel of Speech Act Theory. *Journal of Pragmatics*.

Colin Matheson, Massimo Poesio, and David Traum. 2000. Modelling Grounding and Discourse Obligations Using Update Rules. In *Proceedings of the 1st North American chapter of the Association for Computational Linguistics conference*, pages 1–8. Association for Computational Linguistics.

Vivek K. Rangarajan Sridhar, Srinivas Bangalore, and Shrikanth Narayanan. 2007. Exploiting Prosodic Features for Dialog Act Tagging in a Discriminative Modeling Framework. In *Proceedings of InterSpeech*, pages 150–153, Antwerp, Belgium.

Vivek K. Rangarajan Sridhar, Srinivas Bangalore, and Shrikanth Narayanan. 2009. Combining Lexical, Syntactic and Prosodic Cues for Improved Online Dialog Act Tagging. *Computer Speech & Language*, 23(4):407–422.

François Récanati. 1991. The Pragmatics of What is Said. In Steven Davis, editor, *Pragmatics: a reader*, pages 97–120. Oxford University Press, New York.

Sophie Rosset and Lori Lamel. 2004. Automatic Detection of Dialog Acts Based on Multi-level Information. In *in ICSLP, Jeju Island*, pages 540–543.

Sophie Rosset, Delphine Tribout, and Lori Lamel. 2008. Multi-level information and automatic dialog act detection in human-human spoken dialogs. *Speech Communication*, 50(1):1–13.

M. David Sadek, A. Ferrieux, A. Cozannet, Philippe Bretier, Franck Panaget, and J. Simonin. 1996. Effective Human-Computer Cooperative Spoken Dialogue: the AGS Demonstrator. In *Spoken Language, 1996. ICSLP 96. Proceedings., Fourth International Conference on*, pages 546–549 vol.1.

M. David Sadek, Philippe Bretier, and Franck Panaget. 1997. ARTIMIS: Natural Dialogue Meets Rational Agency. In *in Proceedings of IJCAI-97*, pages 1030–1035. Morgan Kaufmann.

M. David Sadek. 1992. A Study in the Logic of Intention. In *Proceedings of the 3rd International Conference on Principles of Knowledge Representation and Reasoning (KR'92). Cambridge, MA, October 25-29, 1992*, pages 462–473. Morgan Kaufmann.

M. David Sadek. 1994. Communication Theory = Rationality Principles + Communicative Act Models. In *In: Proc. of AAAI 94 Workshop on Planning for Interagent Comm.*

Dan Sperber and Deirdre Wilson. 1995. *Relevance*. Communication and Cognition. Wiley-Blackwell.

Mark Steedman and Jason Baldridge. 2011. Combinatory Categorial Grammar.

Amanda J. Stent. 2000. The Monroe Corpus. Technical report.

Andreas Stolcke, Noah Coccaro, Rebecca Bates, Paul Taylor, Carol Van Ess-Dykema, Klaus Ries, Elizabeth Shriberg, Daniel Jurafsky, Rachel Martin, and Marie Meteer. 2000. Dialogue Act Modeling for Automatic Tagging and Recognition of Conversational Speech. *Computational Linguistics*, 26(3):339–373.

David R. Traum and James F. Allen. 1994. Discourse Obligations in Dialogue Processing. In *Proceedings of the 32nd annual meeting on Association for Computational Linguistics*, pages 1–8. Association for Computational Linguistics.

David R. Traum and Elizabeth A. Hinkelman. 1992. Conversation Acts in Task-oriented Spoken Dialogue. *Computational intelligence*, 8(3):575–599.

David Traum, Johan Bos, Robin Cooper, Staffan Larsson, Ian Lewin, Colin Matheson, and Massimo Poesio. 1999. A Model of Dialogue Moves and Information State Revision. Technical report.

Daniel Couto Vale and Vivien Mast. 2012a. Key Interpersonal Communication Skills for Wheelchairs. In *CogInfoCom '12 Proceedings of the 3rd IEEE International Conference on Cognitive Infocommunications*, pages 421–426, Kosice, Slovakia. IEEE.

Daniel Couto Vale and Vivien Mast. 2012b. Using Foot-Syllable Grammars to Customize Speech Recognizers for Dialogue Systems. In *TSD '12 Lecture Notes in Artificial Intelligence vol. 7499*, Brno, Czech Republic. Springer.